



# Multiplex network inference with sparse tensor decomposition for functional connectivity

Gaetan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves

## ► To cite this version:

Gaetan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves. Multiplex network inference with sparse tensor decomposition for functional connectivity. *IEEE Transactions on Signal and Information Processing over Networks*, 2020, 6, pp.316 - 328. 10.1109/TSIPN.2020.2984853 . hal-02531459

**HAL Id: hal-02531459**

**<https://inria.hal.science/hal-02531459>**

Submitted on 3 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multiplex network inference with sparse tensor decomposition for functional connectivity

Gaëtan Frusque *Student Member, IEEE*, Julien Jung,  
Pierre Borgnat *Member, IEEE*, Paulo Gonçalves *Member, IEEE*.

**Abstract**—Functional connectivity (FC) is a graph-like data structure commonly used by neuroscientists to study the dynamic behaviour of brain activity. However, these analyses rapidly become complex and time-consuming, since the number of connectivity components to be studied is quadratic with the number of electrodes. In this work, we address the problem of clustering FC into relevant ensembles of simultaneously activated components, yielding a multiplex network that reveals characteristic patterns of the epileptic seizures of a given patient. While  $k$ -means is certainly the most popular method for data clustering, it is known to perform poorly on large dimensional data sets, and to be highly sensitive to noise. To overcome the so-called curse of dimensionality, we propose a new tensor decomposition to reduce the size of the data set formed by FC time-series recorded for several seizures, prior to apply  $k$ -means. We propose an adapted procedure to infer a multiplex network from several FC time series, and we emphasise one particular variant that imposes sparsity constraint. Then, we conduct a real case study, applying the proposed sparse tensor decomposition to iEEG data to infer a multiplex network corresponding to the different stages of an epileptic seizure.

**Index Terms**—dynamic networks, graph decomposition, multiplex graph, clustering, dimensionality reduction, sparsity, tensor decompositions, HOSVD, HOOL, functional connectivity, iEEG.

## I. INTRODUCTION

Epilepsy is one of the most common neurological disorders in the world population. About 40% of the patients are drug-resistant, and a surgical operation can be considered to extract the epileptogenic area. To locate this area and understand the evolution of epilepsy, practitioners often use intracranial electroencephalography (iEEG) recordings [1], [2]. The patient stays in the hospital for several days with electrodes implanted inside the brain to record multiple epileptic seizures. The stages of a seizure are distinguished by similar evolutions of the recorded iEEG signals in

different areas of the brain. Functional Connectivities (FC) quantify along time these similarities and they are computed between all pairs of signals, usually employing the spectral coherence or the Phase Locking Value [3]. Considering electrodes as nodes and FCs as weights on the edges, neuroscientists try to infer from the data, relevant graph-like structures evolving through time. The study of the FCs is an expensive and challenging task, with a typical experiment consisting of about 100 electrodes, hence  $\approx 5000$  FC times series. Also, it is generally assumed that the FC dynamics are comparable from one seizure to another for the same patient. Hence, another asset of iEEG monitoring is the possibility to record several seizures. The joint analysis of these records should ease the identification of a dynamical FC pattern common to all seizures, and characteristic of the patient's health disorder. All this calls for a method able to extract, from a FC network inferred from the iEEG signals, what are the relevant sets of simultaneously activated edges.

Having said that, the objective can be framed as the extraction of several sub-graphs characterizing the different spatial networks involved in the epileptic seizures of a patient. This can be seen as an instance of inference of a multiplex network [4], which, associated with the patterns of temporal activation, plays the role of the decomposition of the dynamic network describing the iEEG recording.

More precisely, the goal is to cluster the FC dynamic graph into sub-graphs, using  $k$ -means. However, applied to noisy data – due to spurious edges – that moreover lie in a high dimensional space,  $k$ -means algorithm is known to perform poorly [5]. To face this issue, we develop a dimensionality reduction method that summarizes into factors of lower dimension the relevant features of FCs that occur in all seizures. Then, we perform  $k$ -means clustering on these reduced factors. Since the data correspond to FCs evolving through time and for different seizures, they can be conveniently represented as a 3-modes tensor. The dimensionality reduction we propose is a specific and original tensor decomposition that imposes structural constraints to the solution, including sparsity and time coherence along the different recordings.

Considering dynamic graphs as tensors is not new, and previous works have already explored this idea, e.g. for sampling [6], dynamic graph inference [7] or community detection [8]. The method we are developing is interpretable as a decomposition of dynamical graphs. The data are decomposed into a structural (i.e., sub-graphs of functional

Manuscript submitted May 28, 2019. This work was supported by the ANR-14-CE27-0001 GRAPHSIP grant, and by the ACADEMICS grant of the IDEXLYON, project of the Université de Lyon, PIA operated by ANR-16-IDEX-0005.

G. Frusque and P. Gonçalves are with Univ Lyon, Inria, CNRS, ENS de Lyon, UCB Lyon 1, LIP UMR 5668, F-69342, Lyon, France (e-mail: gaetan.frusque@ens-lyon.fr and paulo.goncalves@inria.fr).

J. Jung is with HCL, Neuro. Hosp., Functional Neurology and Epileptology Dept & Lyon Neurosc. Res. Cent., INSERM, CNRS, Lyon, France (e-mail: julien.jung@chu-lyon.fr).

P. Borgnat is with Univ Lyon, CNRS, ENS de Lyon, UCB Lyon 1, Laboratoire de Physique, UMR 5672, F-69342 Lyon, France (e-mail: Pierre.Borgnat@ens-lyon.fr).

connectivity) and temporal (i.e., time-varying activation) signatures, here of epileptic seizures. This procedure could be used in other contexts involving dynamic graphs, such as social face-to-face interactions or transportation networks.

Our contribution is to develop a new algorithm for dimension reduction of tensor data that imposes relevant structural constraints on the solutions. It addresses the necessity to extract patterns common to all seizures. Then, thanks to this algorithm, we validate a relevant data processing workflow allowing us to infer a decomposition as a multiplex network from iEEG recordings, considered as dynamic graphs. To assess the capacity of the method at compressing data while enhancing structural features, we compare it with other dimension reduction approaches (including other tensor decompositions).  $k$ -means is applied to cluster the reduced factor matrix obtained from each method to find out which reduction provides the best partitioning. Finally, this decomposition of dynamic graphs is applied to real iEEG recordings, to characterise the prime lineament of functional connectivity during epileptic seizures of a patient.

The article is organized as follows. Section II sets the general notations. Section III describes the problem of multiplex networks decomposition, and it reviews the methodological background, from  $k$ -means to tensor decomposition methods. Section IV details the new tensor decomposition we propose. In Section V, we compare on a simple, yet flexible model of iEEG dynamic networks, our method to the state-of-the-art approaches. In the course, we highlight the performance and limitations of the proposed method. Section VI applies the method on real data, exhibiting its capacity to infer a dynamic graph characterizing the time evolution of connectivity during a seizure. We conclude in Section VII.

## II. NOTATIONS

To keep the presentation as clear as possible, we introduce the following notation coming from [9], [10]. We also refer to these works as good introductions to tensor decomposition theory. In [11], a more recent and in-depth introduction is provided. A  $D$ -modes array (where  $D$  corresponds to the number of dimensions used to represent the data) is called a vector if  $D = 1$ , a matrix if  $D = 2$  and a tensor if  $D = 3$  or above. In this work, we only use 3-modes tensors, but the presented theory can be generalized to higher dimensions. Tensors are denoted with bold case calligraphic letters  $\mathcal{X}$ , matrices and vectors are denoted respectively in bold upper-case and lower-case  $\mathbf{X}$ ,  $\mathbf{x}$ , and scalars by lower-case letters  $x$ . The indices  $l$ ,  $t$ ,  $s$ ,  $k$  and  $n$  denote the FC, the time, the epoch (or trial), the factor and the cluster modes respectively, while  $L, T, S, K$  and  $N$  stand for their corresponding upper bounds.  $\mathbf{X}^t$  is the transpose of matrix  $\mathbf{X}$ ;  $\mathbf{x}_{:,t}$  (resp.  $\mathbf{x}_{t,:}$ ), corresponds to the  $t$ -th column (resp. the  $l$ -th row) of matrix  $\mathbf{X} \in \mathbb{R}^{L \times T}$ . The matrices  $\mathbf{X}_{l,:}$ ,  $\mathbf{X}_{:,t}$  and  $\mathbf{X}_{:,s}$  correspond to slices of the tensor  $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$  in each of its three modes, and  $\mathbf{X}_{:,s}$  is referred to as an epoch (or trial).

The matricization corresponds to the matrix representation of a tensor and it can be made for each mode. For a 3-modes tensor they are noted  $\mathbf{X}_{(L)} \in \mathbb{R}^{L \times TS}$  (mode-1 matricization),  $\mathbf{X}_{(T)} \in \mathbb{R}^{T \times LS}$  (mode-2 matricization) and  $\mathbf{X}_{(S)} \in \mathbb{R}^{S \times LT}$  (mode-3 matricization) [10]. Writing  $[\mathbf{A}, \mathbf{B}]$  the concatenation of two matrices  $\mathbf{A}$  and  $\mathbf{B}$  having the same number of rows, the mode-1 matricization reads:

$$\mathbf{X}_{(L)} = [\mathbf{X}_{:,1}, \mathbf{X}_{:,2}, \dots, \mathbf{X}_{:,S}]. \quad (1)$$

Similarly, the vectorisation of a matrix  $\mathbf{X} \in \mathbb{R}^{L \times T}$ , denoted by  $\text{vec}(\mathbf{X}) \in \mathbb{R}^{1 \times LT}$  converts a matrix to a vector [10]:

$$\text{vec}(\mathbf{X}) = [\mathbf{x}_{:,1}^t, \mathbf{x}_{:,2}^t, \dots, \mathbf{x}_{:,L}^t]^t. \quad (2)$$

Considering  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^L$ ,  $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{l=1}^L x_l y_l$  is the scalar product between two vectors and  $\|\mathbf{x}\|_F = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ , the Frobenius norm. These notations can be extended to  $D$ -mode arrays [10], [12], e.g.  $\langle \mathbf{X}, \mathbf{Y} \rangle = \langle \text{vec}(\mathbf{X}), \text{vec}(\mathbf{Y}) \rangle$  or  $\|\mathcal{X}\|_F = \|\text{vec}(\mathbf{X}_{(L)})\|_F$ . The  $L_1$ -norm for a matrix is defined as:

$$\|\mathbf{X}\|_1 = \sum_{l=1}^L \sum_{t=1}^T |x_{lt}|. \quad (3)$$

Finally, as in [11], the Kronecker product between two matrices is noted  $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$  with  $\mathbf{A} \in \mathbb{R}^{L \times T}$ ,  $\mathbf{B} \in \mathbb{R}^{K \times N}$  and  $\mathbf{C} \in \mathbb{R}^{LK \times TN}$ , and the Khatri-Rao product is noted  $\mathbf{E} = \mathbf{A} \odot \mathbf{D}$ , with  $\mathbf{D} \in \mathbb{R}^{K \times T}$  and  $\mathbf{E} \in \mathbb{R}^{LK \times T}$ :

$$\mathbf{C} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1T}\mathbf{B} \\ \vdots & & \vdots \\ a_{L1}\mathbf{B} & \dots & a_{LT}\mathbf{B} \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} a_{11}\mathbf{d}_{:,1} & \dots & a_{1T}\mathbf{d}_{:,T} \\ \vdots & & \vdots \\ a_{L1}\mathbf{d}_{:,1} & \dots & a_{LT}\mathbf{d}_{:,T} \end{pmatrix}. \quad (4) \quad (5)$$

## III. PROBLEM DESCRIPTION AND RELATED METHODS

### A. Multiplex network inference

As discussed in the Introduction, our objective is to infer components that are sub-graphs of dynamic networks. Let us start with a brief description of the pre-processing steps of iEEG recordings. For a given seizure, all recorded signals are chopped into  $T$  segments (possibly overlapping); a functional connectivity is computed between each pair of electrodes and over each segment. From the resulting weighted FC values, we form a dynamic network composed of  $T$  weighted graphs  $\mathcal{G}_t$ , coinciding each, to a (noisy) measurement of the cortical network's state at instant  $t \in \{1, \dots, T\}$ .

As our aim is to find relevant clusters of functional connectivities, i.e. the edges of  $\mathcal{G}_t$ , it is more appropriate to consider the line-graph, rather than the (dynamic) graph directly. We recall that the line-graph, noted  $\mathcal{LG}_t$ , is the graph where the vertices are the edges of  $\mathcal{G}_t$ , and an edge exists between 2 vertices of  $\mathcal{LG}_t$  only if the associated edges in  $\mathcal{G}_t$  share a common vertex. The (upper diagonal) elements of the adjacency matrix of  $\mathcal{G}_t$ , containing the  $L$  values of FCs, corresponds then to a graph signal on  $\mathcal{LG}_t$ . The data matrix  $\mathbf{X} \in \mathbb{R}^{L \times T}$  for one epoch (seizure) corresponds to one such graph signal on  $\{\mathcal{LG}_t\}_{t=1, \dots, T}$ .

The rationale for using line-graph instead of the original graph, is that we cluster FCs that activate synchronously across time. This is different from the clustering at the node-level, which is the usual context for sub-graphs inference (be it for community detection in complex networks [13], [14] or group inference in Stochastic Block Models of graphs [15]–[17]). Such methods tend to extract complete sub-graphs, an hypothesis that is not consistent in the context of our application. Ideally, we aim to identify  $N$  clusters of FCs (and not of nodes) from the Epoch  $\mathbf{X}$ , yielding  $N - 1$  sub-graphs containing edges that activate synchronously across time, and an additional graph gathering all asynchronous FCs. Considering line-graph  $\mathcal{L}\mathcal{G}_t$ , with the FC signals lying on top of it, adds in a conceptual freedom: clustering the nodes of  $\mathcal{L}\mathcal{G}_t$ , according to their signals  $\mathbf{x}_t$ , readily leads to a multiplex network decomposition of  $\mathcal{G}_t$ . An illustration of this decomposition applied to a toy example is presented in Fig. 1. From the signals plotted in Fig. 1(a), we infer the FC matrix  $\mathbf{X}$  displayed in Fig. 1(b). The use of  $k$ -means allows to identify  $N = 3$  clusters in  $\mathbf{X}$ . The three resulting sub-graphs (or layers of the multiplex graph) are represented in Fig. 1(c), where the first two correspond to activation patterns and the third one contains the asynchronous FCs.

In addition, as  $S$  different seizures are observed for the same patient, we stack the resulting epochs in a 3-modes tensor  $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$ . In order to get relevant FC sub-graphs that activate similarly across the different seizures, we need a method able to decompose a tensor structure into multiplex networks. To this end, we first propose to decompose the tensor  $\mathcal{X}$  in three matrices (whose formal definition will be clarified in Section IV), as follows:

$$\mathbf{X}_{(L)} \approx \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t. \quad (6)$$

$\mathbf{X}_{(L)}$  is the  $L$ -mode unfolded tensor matrix.  $\mathbf{F} \in \mathbb{R}^{L \times K}$  is a factor matrix that condenses the relevant FC clusters, with  $K \ll \min(L, T)$ .  $\mathbf{V} \in \mathbb{R}^{T \times K}$  codes for the temporal activation of each factor, and the vector  $\mathbf{w} \in \mathbb{R}^{S \times 1}$  gives the relative contribution of each epoch to the common pattern. By imposing relevant constraints on  $\mathbf{F}$  and  $\mathbf{V}$ , we expect that  $k$ -means applied to the factor matrix  $\mathbf{F}$ , produces the appropriate multiplex network of FCs from the data tensor  $\mathcal{X}$ .

### B. Tensor clustering with $k$ -means

There exist many approaches to cluster FC matrix, using various techniques. For example, community detection [18] discriminates FCs characterising different states of an epileptic seizure. Spectral clustering [19] combines electrodes to cluster complete FC graphs. Non-negative matrix factorisation [20] discovers subgraphs of FC with evolutionary activation over time.  $k$ -means is used to cluster FC or time states in [20]–[22]. The clustering method we use here is the  $k$ -means algorithm [5], because it is simple to apply, and it performs well against other more recent clustering methods [23], like dbSCAN [24] or spectral clustering [25]. Moreover, its limitations are well studied (we will recall them later). Applying a Higher-Order version

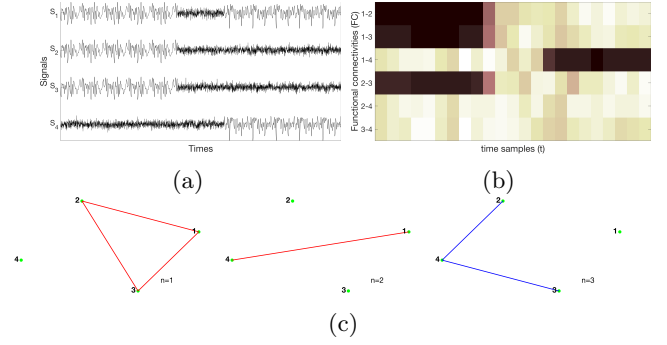


Fig. 1: (a) 4 simulated iEEG signals, (b) Matrix  $\mathbf{X} \in \mathbb{R}^{L \times T}$  of FCs computed between the 6 pairs of signals by PLV (no threshold used), (c) Multiplex graph obtained after applying  $k$ -means on the FC matrix for  $N = 3$  clusters. For the representation, electrodes 1,2,3 and 4 are nodes. The binary edges in each layer of the multiplex graph corresponds to the clustering solution; the two red graph layers have edges that activate synchronously through time, the blue graph layer contains the remaining asynchronous FCs.

of  $k$ -means algorithm [5] on the tensor  $\mathcal{X}$  amounts to find the  $N$  matrices  $\Theta_{n::}$ , centroids of the  $N$  clusters that best characterise the data. The  $l$ -th slice  $\mathbf{X}_{l::}$  of the tensor, belongs to the cluster  $n$  (noted  $l \in C_n$ ), if its nearest centroid is  $\Theta_{n::}$ . Interestingly, and as shown in Appendix A-A, this is equivalent to applying  $k$ -means on the rows of the matrix  $\mathbf{X}_{(L)}$ . As it is a concatenation of all epochs, this matrix allows for a straightforward study of multiple seizures [26]. Notice that in this case, the centroids of the clusters become vectors  $\theta_n \in \mathbb{R}^{TS}$ . Then, clustering a tensor with  $k$ -means simply amounts to solve the following optimization problem:

$$\arg \min_{\theta_1, \theta_2, \dots, \theta_N} \sum_{n=1}^N \sum_{l \in C_n} \|\mathbf{X}_{(L)l:} - \theta_n\|_F^2. \quad (7)$$

However, as  $\mathbf{X}_{(L)}$  lies in a high dimensional space ( $TS \gg 0$ ) and the problem in (7) being known to be NP-hard, performances of  $k$ -means can drastically degrade [5]. Moreover, FC measurements are often noisy due to spurious similarities between iEEG signals. All this induces the existence of local minima that do not necessarily correspond to relevant clustering solutions, and the number of repetitions with different initial conditions that are needed to warrant a satisfactory local minimum jeopardizes the formulation in (7) to treat our application. On the other hand, we can cast  $k$ -means as a problem of constrained matrix factorisation [27], [28]. This remark naturally justifies to perform, beforehand, a dimensionality reduction producing a relaxation of the  $k$ -means solution. As we show it in Appendix A-B, solving Eq. (7) is equivalent to find the solution of the norm maximisation:

$$\arg \max_{\mathbf{A}} \|\mathbf{A}^t \mathbf{X}_{(L)}\|_F^2 \quad (8)$$

where the columns of  $\mathbf{A}$  are constrained to form a standard basis for a subspace of  $\mathbb{R}^L$ . It also means that  $\mathbf{A}$  is unitary, sparse and non-negative. To bypass the curse of

dimensionality, we propose to apply  $k$ -means to a factor matrix  $\mathbf{F} \in \mathbb{R}^{L \times K}$  that is a reduction of the raw data  $\mathbf{X}_{(L)}$ . This yields the surrogate maximisation:

$$\arg \max_{\mathbf{A}} \|\mathbf{A}^t \mathbf{F}\|_F^2. \quad (9)$$

Although the solutions of (9) and of (8) should share the same structure, it is important to notice that in general, they are not equal. Indeed, as a by-product, the dimensionality reduction also attenuates the effect of spurious connectivities by squeezing the data into its factors, easing thus the clustering task. In addition, if the reduction also imposes to  $\mathbf{F}$  constraints that  $\mathbf{A}$  must satisfy, then the clustering solution can be improved [28].

In Section IV, we will propose a decomposition as in Eq. (6) to get a reduced dimensional factor matrix  $\mathbf{F}$ .

### C. Related works on dimensionality reduction

For  $k$ -means, the singular value decomposition (SVD) is often used for dimensionality reduction [29]–[31]. The tandem of SVD and  $k$ -means clustering was applied to network applications [29], [32], in medical imaging [33] or FC analysis [20] and thoroughly studied in gene expression [30]. The denoising properties of the SVD are well understood, limiting the noise to the reduced signal subspace only [34]. Applied on  $\mathbf{X}_{(L)}$ , the SVD reads [31]:

$$\mathbf{X}_{(L)} = \mathbf{U} \mathbf{\Lambda} \mathbf{Z}^t, \quad (10)$$

where  $\mathbf{U} \in \mathbb{R}^{L \times L}$  and  $\mathbf{Z} \in \mathbb{R}^{TS \times TS}$  are unitary matrices, and  $\mathbf{\Lambda} \in \mathbb{R}^{L \times TS}$  is a positive diagonal matrix such that, if  $L \leq TS$ ,  $\lambda_{11} \geq \lambda_{ll} \geq \lambda_{LL}$ . The best low order approximation according to the Frobenius norm of  $\mathbf{X}_{(L)}$  is  $\mathbf{U}^{(K)} \mathbf{\Lambda}^{(K)} \mathbf{Z}^{(K)t}$  with  $\mathbf{U}^{(K)} = \mathbf{U}_{:[1, \dots, K]}$ ,  $\mathbf{Z}^{(K)} = \mathbf{Z}_{:[1, \dots, K]}$  and  $\mathbf{\Lambda}^{(K)} = \mathbf{U}^{(K)t} \mathbf{X} \mathbf{Z}^{(K)}$ . Combining such dimension reduction with  $k$ -means corresponds to solve Eq. (9) when  $\mathbf{F}_{\text{SVD}} = \mathbf{U}^{(K)} \mathbf{\Lambda}^{(K)}$ . Also, recent theoretical results in [35] point that, under some hypotheses,  $\mathbf{F}_{\text{SVD}}$  is a good approximation of  $\mathbf{X}_{(L)}$  to obtain the best clustering from Eq. (8). However, this matrix approach prevents the use of structural constraints on the temporal and seizure modes separately, since they are merged in one single mode. A significant improvement is to consider tensor decompositions to estimate the factor matrix  $\mathbf{F}$ .

Several extensions of SVD to tensors were proposed [9] and applied in the context of brain data, including FC [36]. One of the most famous extension is the high order SVD (HOSVD) [12]. It consists of finding three normalised matrices  $\mathbf{U}^{(K_L)} \in \mathbb{R}^{L \times K_L}$ ,  $\mathbf{V}^{(K_T)} \in \mathbb{R}^{T \times K_T}$  and  $\mathbf{W}^{(K_S)} \in \mathbb{R}^{S \times K_S}$  such that the following decomposition holds:

$$\arg \min_{\mathbf{U}^{(K_L)}, \mathbf{V}^{(K_T)}, \mathbf{W}^{(K_S)}} \|\mathbf{X}_{(L)} - \mathbf{U}^{(K_L)} \mathbf{G}_{(L)} (\mathbf{W}^{(K_S)} \otimes \mathbf{V}^{(K_T)})^t\|_F^2, \quad (11)$$

with  $\mathbf{G}_{(L)} \in \mathbb{R}^{K_L \times K_T K_S}$  a dense matrix. Like SVD, HOSVD seeks for orthogonal matrices that best characterise the data with respect to each of their modes. If  $K_L = L$ ,  $K_T = T$  and  $K_S = S$ , an optimal solution can be directly computed using the HOSVD algorithm

proposed in [12]. However (in contrast to the SVD), the sought low-rank matrices do not stem from truncating the solution  $\mathbf{U}$  (resp.  $\mathbf{V}$ ,  $\mathbf{W}$ ) to its first  $K_L$  (resp.  $K_T$ ,  $K_S$ ) columns. Instead, it is required to resort to an iterative optimization algorithm as the Higher-Order Orthogonal Iteration of Tensors (HOOI) proposed in [37], [38] and detailed in Algorithm 1. The work in [39] shows that in most cases, HOOI reaches the optimal solution of Eq. (11), with good convergence performance. This yields a reduced dimension factor matrix  $\mathbf{F}_{\text{HOOI}} = \mathbf{U}^{(K_L)} \mathbf{G}_{(L)}$  in Eq. (9), on which  $k$ -means can apply.

Note that HOSVD was already used to extract a representative Epoch from a 4-modes FC dataset in [40] and in [41], using a Higher-Order robust principal component analysis. However, both methods suffer from a lack of sparsity in their resulting graphs, entailing results that are difficult to analyse as they do not single out any critical functional connectivity that would characterise the seizure onset. Another extension of the SVD is the canonical polyadic decomposition (or CP, PARAFAC), it was used in [42] to exhibit the principal data patterns. However, this decomposition is limited in our case study, because it does not look for a typical pattern that is common to all seizures.

All that calls for a new tensor decomposition able to exhibit FC factors that are sparse and that are common to all seizures.

---

**Algorithm 1 HOOI.** Estimation of  $\mathbf{U}^{(K_L)}$ ,  $\mathbf{V}^{(K_T)}$  and  $\mathbf{W}^{(K_S)}$ .

---

**Require:**  $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$ , the parameters for reduction ( $K_L, K_T, K_S$ ), increment tolerance  $\epsilon > 0$ ,  $i_{\max}$ .

$i = 0$

$[\mathbf{U}_0, \mathbf{V}_0, \mathbf{W}_0] = \text{HOSVD}(\mathcal{X})$

**while**  $i < i_{\max}$  or  $\|\mathbf{G}_{(L)i}\|_F^2 - \|\mathbf{G}_{(L)i-1}\|_F^2 > \epsilon$  **do**  
 $\mathbf{U}_{i+1}^{(K_L)} = \text{left-}K_L\text{-SVD}(\mathbf{X}_{(L)}(\mathbf{W}_i^{(K_S)} \otimes \mathbf{V}_i^{(K_T)}))$   
 $\mathbf{V}_{i+1}^{(K_T)} = \text{left-}K_T\text{-SVD}(\mathbf{X}_{(T)}(\mathbf{U}_{i+1}^{(K_L)} \otimes \mathbf{W}_i^{(K_S)}))$   
 $\mathbf{W}_{i+1}^{(K_S)} = \text{left-}K_S\text{-SVD}(\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K_T)} \otimes \mathbf{U}_{i+1}^{(K_L)}))$   
 $\mathbf{G}_{(L)i+1} = \mathbf{U}_{i+1}^{(K_L)t} \mathbf{X}_{(L)} (\mathbf{W}_{i+1}^{(K_S)} \otimes \mathbf{V}_{i+1}^{(K_T)})$   
 $i = i+1$

**end while**

---

<sup>1</sup>  $\mathbf{U}^{(K)} = \text{left-}K\text{-SVD}(\mathbf{X})$  refers to the algorithm to compute the first  $K$  singular components of  $X$ .

## IV. NEW TENSOR DECOMPOSITION FOR CLUSTERING

### A. Proposed decomposition

Our goal is to perform a decomposition similar to that of Eq. (6), able to isolate the dynamical FC components appearing in all seizures. Hence, setting  $K_S = 1$  in (11) leads to a vector  $\mathbf{w} \in \mathbb{R}^{S \times 1}$  that measures the contribution rate of each seizure to the common epoch. For instance, a constant vector ( $\mathbf{w}_s = \frac{1}{\sqrt{S}} \forall s \in 1, \dots, S$ ) means that the common pattern is simply the average of all the seizures. In the opposite, if the seizures share no common features, the retained component matches the seizure  $s$  for which  $\|\mathbf{X}_{::s}\|_F^2$  is maximum. Then, the common epoch is a linear

combination of all seizure's patterns, and it is always good practice to have a close look at  $\mathbf{w} \in \mathbb{R}^S$  to assess the heterogeneity of the seizures.

For the sake of simplicity and without loss of generality, we consider in the following that  $K_L = K_T = K$ . We propose a tensor factorisation that constrains the temporal patterns common to all epochs and the factor matrix  $\mathbf{F} \in \mathbb{R}^{L \times K}$  to be both sparse. Regarding the factor matrix, the sparsity constraint is compelling because otherwise,  $\mathbf{F}$  becomes close to a cluster assignment matrix. In addition, it limits the complexity of clusters by reducing the number of FCs they contain. This is important in the context of epileptic data where a large number of FC measurements can be passively implied in a neurological process (during the discharge of the seizure, for example). Additionally, the temporal patterns  $\mathbf{V} \in \mathbb{R}^{T \times K}$  are also forced to be sparse to select specific periods of FC activation and to eliminate periods where there is no common activation of FC clusters. All in all, the decomposition we propose is the solution to the following optimisation problem:

$$\begin{aligned} \underset{\mathbf{U}, \mathbf{V}, \mathbf{w}}{\operatorname{argmin}} \quad & \|\mathbf{X}_{(L)} - \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1 \\ \text{s.t.} \quad & \lambda_2 \|\mathbf{v}_{:,k}\|_1 + \|\mathbf{v}_{:,k}\|_F^2 \leq 1 \\ \text{and} \quad & \mathbf{w}^t \mathbf{w} = 1 \end{aligned} \quad (12)$$

where the meta-parameters  $\lambda_1$  and  $\lambda_2$  allow for tuning the trade-off between accuracy and sparsity of the approximation. It is worthwhile to stress the importance of the elastic-net constraint applied on  $\mathbf{V}$ , rather than just a simple lasso regularisation, as it is discussed in [43]. Likewise, imposing only sparse regularisation on both modes of a tensor decomposition often leads to matrices containing co-linear factors and thus, to trivial solutions. On the contrary, applying an elastic-net constraint prevents the occurrence of such irrelevant decompositions [44]. Note also, that  $\mathbf{F}$  in problem (12), is scaled such that  $\|\mathbf{w}\|_F^2 = \|\mathbf{v}_{:,k}\|_F^2 = 1, \forall k$ . By smoothing out the scaling indeterminacy, we transfer all the energy to the factor matrix  $\mathbf{F}$ . We call the decomposition in (12), the Higher-Order Sparse Mode Reduction (HOSMR). Its optimal solution yields a low dimensional factor matrix  $\mathbf{F}_{\text{HOSMR}}$  that we plug in Eq. (9) to perform  $k$ -means clustering.

### B. Relating HOSMR to other sparse tensor decompositions

We can relate the HOSMR loss function in (12), to already existing sparse tensor decompositions. First, the proposed decomposition is close to a Block term decomposition [45] with one component, and with an additional sparsity constraint imposed on the two first modes. It is also a variant of the parsimonious Tucker decomposition presented in [46], [47]. But the main difference in our proposition is that we are not interested in a core tensor, which would not be compatible with the constraint of sparse temporal activation profiles of sparse FC patterns. Instead, we integrate the core tensor into the factor matrix  $\mathbf{F}$ . Then,

denoting  $f$  the HOSVD loss function of Eq. (11), when  $K_S = 1$ ,  $K_L = K_T = K$ , we get:

$$f(\mathbf{U}, \mathbf{V}, \mathbf{w}) = \|\mathbf{X}_{(L)} - \mathbf{U}\mathbf{G}_{(L)}(\mathbf{w} \otimes \mathbf{V})^t\|_F^2. \quad (13)$$

To identify temporal activation along with the sparse FC factors, we apply a lasso regularisation to the factor matrix  $\mathbf{U}\mathbf{G}_{(L)}$ . It leads us to the regularized loss function:

$$f^{\text{reg}}(\mathbf{U}, \mathbf{V}, \mathbf{w}) = f(\mathbf{U}, \mathbf{V}, \mathbf{w}) + \lambda_1 \|\mathbf{U}\mathbf{G}_{(L)}\|_1. \quad (14)$$

Then, for the sake of simplifying the loss function, and keeping in mind that our interest lies in the factor matrix, and not in the analysis of the core tensor nor in that of  $\mathbf{U}$ , we combine these two in one single matrix  $\mathbf{F} = \mathbf{U}\mathbf{G}_{(L)} \in \mathbb{R}^{L \times K}$ . The HOSMR decomposition we propose in Eq. (12) follows, with a loss function that takes on the form:

$$g(\mathbf{U}, \mathbf{V}, \mathbf{w}) = \|\mathbf{X}_{(L)} - \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1. \quad (15)$$

Different implementations to perform sparse tensor decompositions exist. For instance, in [46], an algorithm is proposed to perform parsimonious Tucker decomposition. However, it consists of adding several lasso regularisations to the sought modes and, as we already mentioned, this is not appropriate when we expect sparsity in more than one mode [43]. A similar decomposition using elastic net regularisations on each mode is proposed in [47]. However, factorisation with a core tensor impairs the interpretability of the components, and the association between the FC patterns and the activation profiles breaks down. As for [48], the authors use a power method, which is known not to provide the best solution [9], and to lack interpretability of its components. As it is done in [37], and [48], we could also impose orthogonality and sparsity jointly, rather than a elastic-net constraint. Yet, the corresponding greedy solution remains challenging to interpret.

Regarding efficiency too, compared to the methods presented above, we pretend that our algorithm converges more rapidly to a stable solution. Indeed, as shown below, the proposed implementation alternates only two steps, instead of three for the competitors.

On the other hand, our method is less general since it is specifically designed for the case  $K_S = 1$ .

### C. Adapted HOOI algorithm restricted to $K_S = 1$

This section shows, in the particular case  $K_L = K_T = K$  and  $K_S = 1$ , how a modification of the HOOI algorithm can expose a matrix factorisation problem. Beyond proposing a novel approach to minimise the loss function of Eq. (13) with a lower complexity than HOOI, we can easily add constraints to the matrix factorisation, and get a straightforward method to solve the problem of Eq. (12).

The HOOI algorithm 1 minimizes the loss function of Eq. (13), with  $\mathbf{U}$ ,  $\mathbf{V}$  and  $\mathbf{w}$  constrained to be semi-orthonormal matrices and  $\mathbf{G}_{(L)} = \mathbf{U}^t \mathbf{X}_{(L)} (\mathbf{w} \otimes \mathbf{V})$ . The matrices  $\mathbf{U}_i$  and  $\mathbf{V}_i$  are computed by extracting the left singular components of two different matrices. Noting that the generic Epoch (associated to  $\mathbf{w}_i$ ) is a matrix, the idea behind the modification is to extract the left and the right

singular components of this matrix to obtain  $\mathbf{U}_i$  and  $\mathbf{V}_i$ . Then, the three steps of the original algorithm reduce to only two. The proposed modified HOOI (termed mHOOI) alternates the following 2 steps procedure:

(A) *Factorisation of the generic Epoch*: At iteration  $(i + 1)$ , we assume  $\mathbf{w}_i$  to be known; thus, as shown in Appendix B-A, the loss function  $f(\mathbf{U}, \mathbf{V}, \mathbf{w}_i)$  can be rearranged to get the following optimisation criteria:

$$\begin{aligned} & \underset{\mathbf{U}, \mathbf{V}}{\operatorname{argmax}} \|\mathbf{U}^t \mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T) \mathbf{V}\|_F^2, \\ & \text{s.t. } \mathbf{U}^t \mathbf{U} = \mathbf{I}_L, \quad \mathbf{V}^t \mathbf{V} = \mathbf{I}_T. \end{aligned} \quad (16)$$

$\mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T) \in \mathbb{R}^{L \times T}$  is the generic Epoch according to the vector  $\mathbf{w}_i$  (it corresponds to the mode- $L$  matricization of the contraction product between the tensor  $\mathcal{X}$  and the vector  $\mathbf{w}^t$  [10]).  $\mathbf{I}_T \in \mathbb{R}^{T \times T}$  is the identity matrix. The optimal solutions of Eq. (16) are the low-rank approximations  $\mathbf{U}_{i+1}$  and  $\mathbf{V}_{i+1}$ , obtained as the  $K$  first components of the singular value decomposition of  $\mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T) = \mathbf{U}_{i+1} \mathbf{\Lambda}_{i+1} \mathbf{V}_{i+1}^t$ . We then have  $f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) \leq f(\mathbf{U}_i, \mathbf{V}_i, \mathbf{w}_i)$ .

(B) *Update of seizures' contribution*: In order to update  $\mathbf{w}_{i+1}$ , we perform the same step as in HOOI algorithm (Algo 1). We filter the tensor  $\mathcal{X}$  by projecting it on the subspaces spanned by the matrices  $\mathbf{U}_{i+1}$  and  $\mathbf{V}_{i+1}$  and then, we compute the dominant subspace in the seizure mode. The filtering and the decomposition operations used to get  $\mathbf{w}_{i+1}$ , reduce to the problem of computing the dominant left singular vector of  $\mathbf{X}_{(S)}(\mathbf{V}_{i+1} \otimes \mathbf{U}_{i+1})$ . The optimal solution  $\mathbf{w}_{i+1}$  minimizes the loss function  $f(\mathbf{U}, \mathbf{V}, \mathbf{w})$  with  $\mathbf{U} = \mathbf{U}_{i+1}$ ,  $\mathbf{V} = \mathbf{V}_{i+1}$ , and with the constraint  $\mathbf{w}^t \mathbf{w} = 1$ . As demonstrated in appendix (B-A), it corresponds to solve:

$$\begin{aligned} & \underset{\mathbf{w}}{\operatorname{argmax}} \|\mathbf{w}^t \mathbf{X}_{(L)}(\mathbf{V}_{i+1} \otimes \mathbf{U}_{i+1})\|_F^2. \\ & \text{s.t. } \mathbf{w}^t \mathbf{w} = 1, \end{aligned} \quad (17)$$

which in turn, implies that :

$$f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i). \quad (18)$$

Algorithm 2 summarizes the mHOOI method we propose, and shows how to derive the corresponding lower dimension factor matrix  $\mathbf{F}_{mHOOI}$ . It is worth noticing that the stopping criterion we use for the iterative procedure is an angular metric between two successive estimates of  $\mathbf{w}_i$ , which saves the computation of the coefficients  $\mathbf{G}_{(L)i}$ . Algorithm. (2) is an alternating least square algorithm, where we can easily prove the convergence of the loss function since at each step:

$$f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq f(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) \leq f(\mathbf{U}_i, \mathbf{V}_i, \mathbf{w}_i). \quad (19)$$

Compared to the HOOI algorithm, Algorithm (2) entails only two steps instead of three (of similar complexity), and as a result, it should converge faster. Empirical experiments support this assumption (see Appendix B-D, or [44] for more experiments). Finally, the HOOI Algorithm, with  $K_L = K_T = K$  and  $K_S = 1$ , has at each iteration

---

**Algorithm 2 mHOOI.** Estimation of  $\mathbf{F}_{mHOOI}$ 


---

**Require:**  $\mathcal{X}$ ,  $K$ , angular tolerance  $\epsilon > 0$ , et  $i_{max}$ .

```

i = 1
w0 = left-1-SVD(X(S))
while i < i_max or acos(< w_i, w_{i-1} >) > epsilon do
  (A). [U_{i+1}^{(K)}, Lambda_{i+1}, V_{i+1}^{(K)}] = K-SVD(X_{(L)}(w_i \otimes I_T))
  (B). w_{i+1} = left-1-SVD(X_{(S)}(V_{i+1}^{(K)} \otimes U_{i+1}^{(K)}))
  i = i + 1
end while
F_{mHOOI} = U_i^{(K)} Lambda_i

```

---

a computational cost of  $4(2K + 1)LTS$  according to [49]. Since both algorithms perform several SVD, the computational cost of Algorithm (2) is estimated in the same way, and reduces to  $4(K + 1)LTS$  operations at each iteration. Let us emphasise that this decomposition is limited to the case  $K_S = 1$ . When it is not the case, the tensor  $\mathcal{X}$  is compressed into  $K_S$  Epochs that form a tensor. The respective mode- $L$  and mode- $T$  matricization are two matrices with different dimensions, which in turn makes it impossible to extract  $\mathbf{U}_i$  and  $\mathbf{V}_i$  from the same matrix factorisation.

#### D. HOSMR Algorithm

We can now expose our final HOSMR Algorithm that aims at finding a solution to the optimisation problem of Eq. (12). It follows the same rationale as the mHOOI Algorithm (2), but where a sparse matrix factorisation replaces the computation of the SVD of the generic Epoch. The main advantage now, is that we impose a sparse regularisation on  $\mathbf{F}$ , and an elastic-net constraint on  $\mathbf{V}$ , simultaneously. Our Algorithm iterates the two steps:

(A) *Sparse matrix factorisation of the common Epoch*: At iteration  $(i+1)$ , we assume  $\mathbf{w}_i$  to be known. By performing some modifications on the loss function  $g(\mathbf{F}, \mathbf{V}, \mathbf{w}_i)$ , it is shown in Appendix (B-B) that the problem of Eq. (12) can be written as:

$$\begin{aligned} & \underset{\mathbf{F}, \mathbf{V}}{\operatorname{argmin}} \|\mathbf{X}_{(S)}(\mathbf{w}_i \otimes \mathbf{I}_T) - \mathbf{F} \mathbf{V}^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \\ & \text{s.t. } \lambda_2 \|\mathbf{v}_{:k}\|_1 + \|\mathbf{v}_{:k}\|_F^2 \leq 1. \end{aligned} \quad (20)$$

To solve it, we use the low-rank approximation  $\mathbf{F}_{i+1}$  and  $\mathbf{V}_{i+1}$ , i.e. the components of a sparse SVD (or alternately of a sparse PCA) that impose  $\mathbf{F}$  and  $\mathbf{V}$  to be both sparse. Different solutions to perform sparse SVD were proposed in the literature, [50], [51]. In general, both sparsity and orthogonality constraints are imposed to the two components. Here, we use the relaxed version of sparse-SVD proposed in [50], which is more widely preferred in many studies [48], [43]. Under mild conditions, [50] proposes a proof of convergence of the sparse-SVD algorithm, ensuring that  $g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) < g(\mathbf{F}_i, \mathbf{V}_i, \mathbf{w}_i)$ . We initialise the sparse-SVD algorithm using the previous iteration  $\mathbf{F}_i$  and  $\mathbf{V}_i$ . After performing the minimisation, we normalise the columns of  $\mathbf{F}$  and of  $\mathbf{V}$  such that  $\|\mathbf{v}_{:k}\|_F^2 = 1 \forall k \in 1, \dots, K$  in order to remove the scaling indeterminacies between both matrices.

(B) *Update contribution of the seizures:* We assume  $\mathbf{F}_{i+1}$  and  $\mathbf{V}_{i+1}$  to be known. As we show in Appendix (B-C), we can perform a slight modification to the loss function  $g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w})$ , so that the criterion in Eq. (12) becomes:

$$\begin{aligned} \underset{\mathbf{w}}{\operatorname{argmin}} \quad & \|\mathbf{X}_{(S)} - \mathbf{w}\mathbb{1}^t(\mathbf{V} \odot \mathbf{F})^t\|_F^2 \\ \text{s.t.} \quad & \mathbf{w}^t \mathbf{w} = 1 \end{aligned} \quad (21)$$

where  $\mathbb{1} \in \mathbb{R}^K$  is a vector such that  $\mathbb{1}_k = 1, \forall k$ . Eq. (21) corresponds to a least square regression under unit sphere constraint, which is also an instance of the Ridge regression [52]. Calling  $\mathbf{p}^t = \mathbb{1}^t(\mathbf{V} \odot \mathbf{F})^t$ , the Lagrangian function associated to (21) reads:

$$L(\mathbf{w}, \gamma) = \mathbf{w}\mathbf{p}^t\mathbf{p}^t\mathbf{w} - 2\mathbf{w}\mathbf{p}^t\mathbf{X}_{(S)}^t + \gamma(\mathbf{w}^t\mathbf{w} - 1). \quad (22)$$

Differentiating (22) with respect to  $\mathbf{w}$  and  $\gamma$ , leads to the following conditions that the solution  $\mathbf{w}_{i+1}$  must verify

$$\mathbf{w}_{i+1} = \mathbf{X}_{(S)}(\mathbf{p}^t\mathbf{p} + \gamma)\mathbf{p} = \bar{\gamma}\mathbf{X}_{(S)}\mathbf{p} \quad (23)$$

$$\mathbf{w}_{i+1}^t \mathbf{w}_{i+1} = 1 \quad (24)$$

Since  $\mathbf{X}_{(S)}\mathbf{p} \in \mathbb{R}^S$  is a vector,  $\mathbf{w}_{i+1}$  must span the same subspace and the constraint of Eq. (24) imposes to find  $\bar{\gamma} = (\mathbf{p}^t\mathbf{p} + \gamma)$  such that  $\mathbf{w}_{i+1}$  is unit norm:

$$\mathbf{w}_{i+1} = \frac{\mathbf{X}_{(S)}(\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1})\mathbb{1}}{\|\mathbf{X}_{(S)}(\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1})\mathbb{1}\|_F}, \quad (25)$$

Here again, one can prove that :

$$g(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq g(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_i) \quad (26)$$

The whole procedure is summarised in Algorithm 3 whose output is the factor matrix  $\mathbf{F}_{HOSMR}$ . The non-increasing loss function after each iteration:

$$g(\mathbf{U}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}_{i+1}) \leq g(\mathbf{U}_i, \mathbf{V}_i, \mathbf{w}_i) \quad (27)$$

guarantees that Algorithm 3 converges towards a solution.

Following the lines of [50], we can estimate the computational cost of step (A) to be in  $O(L^2TK\rho)$ , where  $\rho$  is the maximal number of iterations allowed to the sparse-SVD algorithm. As for the computational cost of step (B), it shows negligible compared to the one of step (A). Therefore, each iteration of our HOSMR Algorithm has a complexity of  $O(L^2TK\rho)$  and empirical experiments show that it converges in few iterations, typically less than ten.

---

#### Algorithm 3 HOSMR. Estimation of $\mathbf{F}_{HOSMR}$

---

**Require:**  $\mathcal{X}$ ,  $K$ ,  $\lambda_1$ ,  $\lambda_2$ , angular tolerance  $\epsilon > 0$ , et  $i_{max}$ .

$i = 0$

$\mathbf{w}_0 = \text{left-1-SVD}(\mathbf{X}_{(S)})$

**while**  $i < i_{max}$  or  $\text{acos}(\langle \mathbf{w}_i, \mathbf{w}_{i-1} \rangle) > \epsilon$  **do**

(A).  $\mathbf{F}_{i+1}, \mathbf{V}_{i+1}$  minimising (20) with  $\mathbf{X}_{(L)}(\mathbf{w}_i \otimes \mathbf{I}_T)$ .

$\alpha_k = \|\mathbf{v}_{:k}\|_F^2$ ,  $\mathbf{v}_{:k} = \frac{1}{\alpha_k} \mathbf{v}_{:k}$ ,  $\mathbf{f}_{:k} = \alpha_k \mathbf{f}_{:k}$

(B).  $\mathbf{w}_{i+1} = \mathbf{X}_{(S)}(\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1})\mathbb{1} / \|\mathbf{X}_{(S)}(\mathbf{V}_{i+1} \odot \mathbf{F}_{i+1})\mathbb{1}\|_F$

$i = i + 1$

**end while**

$\mathbf{F}_{HOSMR} = \mathbf{F}_i$

---

## V. NUMERICAL EXPERIMENTS ON A FC MODEL

In this section, we evaluate the performance of our proposed tensor decomposition to perform multiplex network inference. We compare our approach with state of the art dimensionality reduction methods. To this end, and because we are primarily interested in iEEG signals, we first present an original and oversimplified model of functional connectivity. Although it is quite elementary, our FC model integrates four sources of uncertainty that enable to reproduce most experimental variabilities.

### A. A dynamic graph model for FC

Epileptic seizure implies a pathological FC that often starts in a focal brain onset, then spreads to the other connected regions, and sometimes split to give rise to new FC components. Fig. 2 displays a characteristic example of actual FC time series measured by the phase lock value (PLV) [3]. The model we propose is aimed at reproducing the global structured patterns of FCs' activation, and the uncertainties of the measures. More precisely, we consider a matrix  $\mathbf{X} \in \mathbb{R}^{L \times T}$ , where  $x_{lt}$  is set to a high value if the FC of index  $l \in \{1, \dots, L\}$  is active at time  $t \in \{1, \dots, T\}$ , and to a low value otherwise (to account for the non ON-OFF discrepancy of the PLV measurement and to the absence of thresholding). We choose values equal to 0.7 and 0.2, respectively. A cluster  $C_n$ ,  $n \in \{1, \dots, N\}$ , is composed of all FCs that are activated over the same period of time  $T_n$ . This is for the deterministic part of the model, defining the structural and temporal pattern of FCs activation, common to all seizures. Superimposed to it, we add a seizure dependent random component, with four uncertainty sources:

- i Random duration: each activation period  $T_n$  is uniformly distributed between a minimum duration (here, 5 time steps) and  $\frac{T}{N}$  time steps. A switch  $\alpha$  allows to toggle between deterministic ( $\alpha = 0$ ) and random ( $\alpha = 1$ ) modes.
- ii Activation error: with probability  $\beta \in [0, 1]$ , each FC of a given group  $C_n$  incurs the risk to be replaced by any other randomly chosen FC.
- iii Connectivity noise: we add to  $x_{lt}$  a white centered Gaussian noise of power  $\sigma^2$  ( $\text{SNR} = \sigma^{-2}$ ).
- iv Jitter: All FCs of the same group start activating with independent jitters, uniformly distributed in  $[-\frac{\delta}{2}, \frac{\delta}{2}]$ .

Figure 2 displays one realisation of these synthetic FC time series. Compared to real data, the model succeeds in reproducing a realistic global pattern. More importantly, it allows to control the nature and the intensity of the variability between the epochs. This serves to evaluate the performance of the dimension reduction methods for clustering, with respect to each source of uncertainty.

We simulate different FC representing  $S$  seizures of a same patient, as i.i.d. realisations of our model with a fixed set of parameters  $\mathbf{b} = [\alpha, \beta, \sigma, \delta]$ . Like for real data, the  $S$  modeled seizures are stacked in a tensor  $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$ . Figure 3 displays the unfolded matrix  $\mathbf{X}_{(L)}$  corresponding to a particular choice of  $\mathbf{b}$ ,  $T$  and  $S$ .



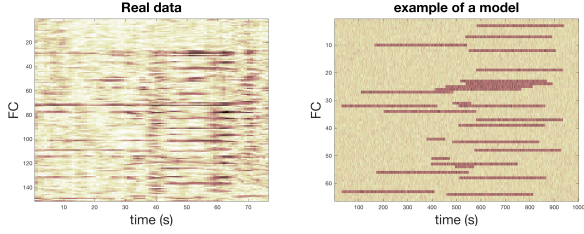


Fig. 2: Real data vs example of model we can get for one seizure with noise parameters  $[\alpha = 1, \beta = 0.2, \sigma = 0.05, \delta = 0]$ .

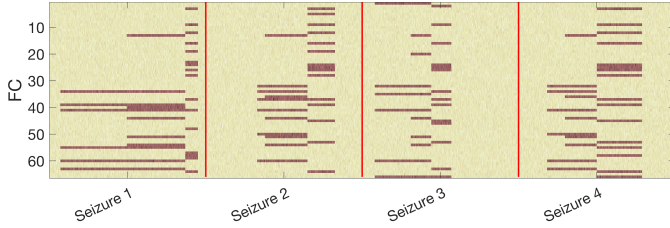


Fig. 3: Realisation of  $\mathbf{X}_{(L)}$  corresponding to  $S = 4$  simulated seizures (or epochs) with noise parameters  $[\alpha = 1, \beta = 0.2, \sigma = 0.05, \delta = 0]$ .

### B. Experimental setup

Using this dynamical graph model, we can now compare the HOSMR method for dimension reduction derived in Section IV, to state-of-the-art methods. Each method yields a factor matrix  $\mathbf{F}$  that serves as an input to  $k$ -means clustering. Performance refer here to the ability at retrieving the FC clusters of the generic pattern. In Table I, we compare 8 methods, and we indicate for each of them, what implementation we used and the best (empirically determined) hyper-parameters values. As benchmarks, we consider the two straightforward approaches that consist in applying  $k$ -means directly on the unfolded tensor  $\mathbf{F}_{Direct} = X_{(L)}$ , or on the seizures' average  $\mathbf{F}_{mean} = 1/S \sum_{s=1}^S X_{::s}$ . The other methods are : (NMF) Non-negative matrix factorisation applied to  $\mathbf{X}_{(L)}$  ; (nnCP) non-negative canonical polyadic decomposition of tensor  $\mathcal{X}$  (see Appendix C for details) ; (sSVD) sparse-SVD applied to  $\mathbf{X}_{(L)}$ ; (mHOOI) modified Higher-Order Orthogonal Iteration of Tensors and (HOSMR) the High Order Sparse Mode Reduction method we propose both applied to  $\mathcal{X}$ .

The reduced dimension  $K$  varies from 1 to 4, and for fair comparison, we retain the value yielding the best grouping score for each method. Regarding  $k$ -means algorithm, we used  $k$ -means++ version [56] that we limited to 1000 iterations and repeated 120 times with different seeds. The number of sought groups is set to  $N = 4$ . Finally, to assess the clustering performance of each method, we use the Adjusted Rand Index (ARI) score [57]. Computed between the resulting multiplex network and the ground truth, this score equals 1 when the match is perfect and 0 if the correspondence does not outperform a mere random clustering.

Method	Section	Implementation	Parameters
1 - Direct	V-B	(1)	-
2 - Mean	V-B	(1)	-
3 - NMF	C	(1)	$\epsilon = 10^{-4}$
4 - nnCP	C	(2)	$\epsilon = 10^{-6}$
5 - sSVD	III-C	(1)	-
6 - mHOOI	IV-C	(1)-(3)	$\epsilon = 10^{-3}$
7 - sSVD	IV-D	(1)-(4)	$\lambda_1 = 0.1, \lambda_2 = 1,$ $\epsilon = 10^{-3}$
8 - HOSMR	IV-D	(1)-(3)-(4)	$\lambda_1 = 0.1, \lambda_2 = 4,$ $\epsilon = 10^{-3}$

TABLE I: List of the compared methods to convert the tensor  $\mathcal{X}$  to a factor matrix  $\mathbf{F}$ . The implementations used for the experiments can be found following the links: 1) Algorithms used in this article: FCTensDec ; (2) N-way toolbox version 3.30 [53] ; (3) MATLAB Tensor Toolbox Version 2.6 [54] ; (4) SPAMS toolbox version 2.6 [55]

### C. Results and discussion

We evaluate the different methods for various uncertainty configurations depicted in Table II. Figures 4 and 5 display the corresponding ARI scores, depending on whether the jitter uncertainty is disabled or enabled. We use error-bar plots to present the results, with the median and the 25th and 75th percentiles estimated out of 120 independent realizations of  $\mathcal{X}$ . We do not show the results when random duration is the sole source of variability, since all methods perform equally well in this case.

Fig. 4 shows that for almost all configurations, methods implying dimension reduction outperform the standard ones. Tensor-based methods perform better than their matrix counterparts, and sparse methods generally outpace conventional methods. On the opposite, the non-negativity constraint does not seem to be helpful here. HOSMR seems to systematically stand out from the other approaches, noticeably for low SNRs. This is also particularly true when connectivity noise is the only source of uncertainty (experiment  $\mathbf{b}_1$ ).

We isolated in Fig. 5 the impact of the jitter  $\delta$ , since it is the only uncertainty source for which HOSMR does not systematically reach the best performance. The worse cases correspond to the combination of jitter and random duration ( $\mathbf{b}_9$ ) and jitter alone ( $\mathbf{b}_7$ ), where our method performs slightly better than the direct methods, but less well than the other methods with dimension reduction. For those configurations, it is very likely that a tensor decomposition with  $K_S > 1$  could produce better results. Indeed, the complexity introduced by jitter, disrupts the synchronous activations of FC within a cluster and leads to a mixing effect between groups. A possible way out, valid for  $K_S > 1$ , is to perform a sparse SVD on the temporal mode and a sparse SVD on the FC mode, separately. This procedure comes with a price, since it no longer amounts to minimising some cost function, and we did not pursue further this study.

Finally, HOSMR dominates again in the presence of connectivity noise, showing that fully connected (weighted) networks do not severely hinder the method.

$\mathbf{b}_i = [\alpha, \beta, \sigma, \delta]$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\alpha$			✓		✓	✓			✓			✓	✓	✓
$\beta$		✓	✓	✓		✓		✓			✓	✓		✓
$\sigma$	✓			✓	✓	✓				✓	✓		✓	✓
$\delta$							✓	✓	✓	✓	✓	✓	✓	✓

TABLE II: A checkmark indicates when the uncertainty in configuration  $\mathbf{b}_i$  is present

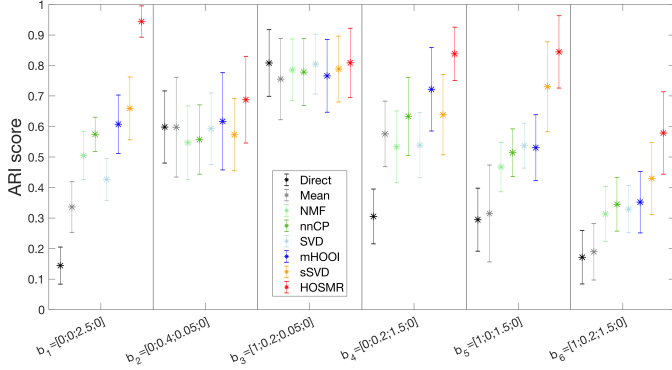


Fig. 4: ARI score for each method, for 6 models associated to vector  $\mathbf{b}_i$  without jitter uncertainty

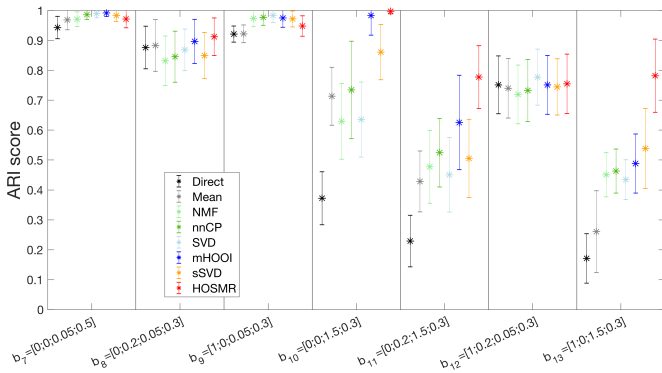


Fig. 5: ARI score for each method, for 7 models associated to vector  $\mathbf{b}_i$  with jitter uncertainty

## VI. APPLICATION ON REAL DATA

**Data.** We now consider real iEEG data from a patient with focal epilepsy [1], [2]. The brain activity is recorded by a tens of stems implanted in the brain's patient. Each stem is composed of 5 to 10 electrodes regularly aligned and spaced by 3.5 mm. During a recording period of 15 days, 4 seizures were observed and isolated. Each seizure is delimited in time by a window of 100 seconds centred on the beginning of the seizures. The signal is sampled at 256 Hz. Among the 108 available electrodes, we retain only 33 uniformly distributed contacts, to avoid too strong spatial correlations. The functional connectivity metric we used is the Phase Locking Value (PLV) [3], and we apply no thresholding to the PLV values, meaning that is very likely to get a fully connected network. The resulting 528 FCs were calculated over a sliding rectangular window of 4 seconds duration, with a time step of one second. After eliminating the points that suffer from border effects, the data is formatted as a tensor  $\mathcal{X} \in \mathbb{R}^{528 \times 96 \times 4}$ .

**Application.** We apply our HOSMR procedure (Algorithm 3) to the tensor  $\mathcal{X}$  to infer the low dimension matrix  $\mathbf{F}_{HOSMR}$ . We empirically set the parameters:  $K = 4$ , in accordance with the neurologists who consider that a seizure is commonly composed of 3 to 5 different steps;  $\gamma_1 = 1$ ,  $\gamma_2 = 0.95$  and  $\epsilon = 10^{-3}$ , correspond to their highest values leading to consistent results.

Fig. 6 (top) shows the temporal activation profiles  $\mathbf{V}$  from Algorithm 3, of the components of  $\mathbf{F}_{HOSMR}$ . There are 4 activation periods, each defining a layer of the multiplex network decomposition. They are associated to the 4 steps of the seizures: before onset, start, propagation and end of the seizure. The time interval around time  $t = 50$  s. is particularly interesting as it shows no coherently activated FC. This is likely to correspond to a functional decoupling [58] at the early start of the seizure, a short period when iEEG activities in different areas of the brain are suddenly decorrelated.

Then, we apply  $k$ -means to  $\mathbf{F}_{HOSMR}$  to identify the  $N = 5$  sub-graphs (FC groups) of the multiplex network decomposition. Since  $\mathbf{F}_{HOSMR}$  is close to an assignment matrix (i.e. a solution  $\mathbf{A}^*$  of Eq. (9)), we match the 4 groups of smaller sizes with the 4 activation periods, whereas the 5th group gathers all remaining asynchronous FCs. Fig. 6 (bottom) shows the positions of the 33 electrodes projected on the transverse plane (according to the Tailarach space). We represent the sub-graphs of FCs thus obtained, by links between the corresponding pairs of electrodes. The four sub-graphs can be associated with four snapshots of a time-varying graph: before the seizure, few electrodes interact in what could correspond to the epileptogenic zone. At the beginning of the seizure, we see a spreading of FC activation with the appearance of a cluster of FCs localised around the epileptogenic zone. During the seizure, other FCs spontaneously appear in the other hemisphere, while the onset FCs diffuse to the left hemisphere. At the end of the seizures, the two hemispheres interact through a global FC sub-graph.

The multiplex network we were able to infer, is in good agreement with the clinical expertise. The focal zone where the seizure originates, is very close to the graphs that activates "before-seizure" and at "seizure start". The propagation of the seizure activity to the right hemisphere of the brain is also well represented by the sub-graphs "Propagation" and "Seizure end".

## VII. CONCLUSION

In this work, we presented different methods to reduce the data dimension before applying  $k$ -means for clustering. Our application target is to identify a time-varying series of functional connectivity that matches the dynamics of epileptic seizures, and that is generic to several epochs. We proposed an original manner to decompose a dynamic, weighted, possibly fully connected, graph into a multiplex network, where each layer corresponds to a set of activated functional connectivities. As we expect all epochs to follow the same storyboard of FC activation, but with variable temporal durations, we developed a tensor decomposition

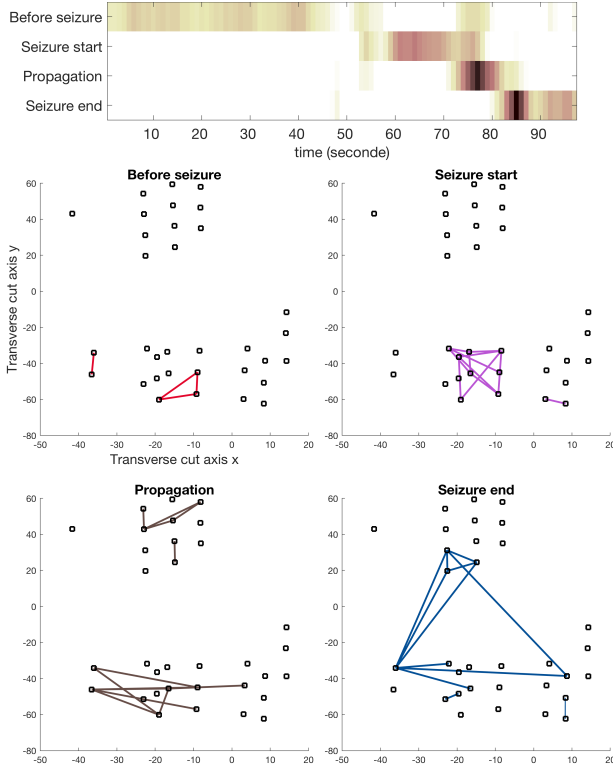


Fig. 6: (top) activation profile of HOSMR (variable  $\mathbf{V}$  of algorithm 3 (bottom) Cluster of FC corresponding to the 4 activation steps of the seizures of the considered patient )

based on existing HOOI algorithms, but capable to select groups of FCs representative of all trials.

We compared the proposed HOSMR method to the state of the art, on a simple, yet flexible mathematical model of epileptic functional connectivity. We obtained clear evidence that in this context, HOSMR outperforms most existing approaches. HOSMR is an exciting alternative to sparse tensor decompositions, that produces a reduced dimensional data set, which can be more efficiently clustered with  $k$ -means.

Applied to real iEEG data recorded during epileptic seizures, HOSMR led us to easily identify the 4 FC activation groups corresponding to the 4 significant periods of seizures' evolution.

As an interesting perspective of this work, we envision to work out an automated selection of the hyper-parameters to permit a more systematic and exhaustive performance comparison. This should also allow for a more in-depth analysis of the functional connectivity and its dynamics, on a wider clinical dataset.

## APPENDIX A

### MORE ABOUT K-MEANS CLUSTERING

#### A. $K$ -means on tensor

The goal here is to perform  $k$ -means to cluster the first mode of the tensor  $\mathcal{X} \in \mathbb{R}^{L \times T \times S}$ . This lead to find  $N$

matrix centroids  $\Theta_{n::} \in \mathbb{R}^{T \times S}$  to cluster the  $L$  slices  $\mathbf{X}_{l::}$ .

$$\underset{\Theta_{1::}, \Theta_{2::}, \dots, \Theta_{N::}}{\operatorname{argmin}} \sum_{n=1}^N \sum_{l \in C_n} \|\mathbf{X}_{l::} - \Theta_{n::}\|_F^2 \quad (28)$$

moreover, the  $L_2$ -norm is invariant the linear function  $\operatorname{vec}(\bullet)$  ( $\|\mathbf{X}\|_F^2 = \|\operatorname{vec}(\mathbf{X})\|_F^2$ ), then :

$$\underset{\Theta_{1::}, \Theta_{2::}, \dots, \Theta_{N::}}{\operatorname{argmin}} \sum_{n=1}^N \sum_{l \in C_n} \|\operatorname{vec}(\mathbf{X}_{l::}) - \operatorname{vec}(\Theta_{n::})\|_F^2 \quad (29)$$

By considering mode-1 matricization of  $\mathcal{X}$  with  $\mathbf{X}_{(1)l} = \operatorname{vec}(\mathbf{X}_{l::})$  and noting  $\theta_n = \operatorname{vec}(\Theta_{n::})$  we get Eq. (7).

#### B. $K$ -means is matrix factorisation

We start from Eq. (7) and set  $\mathbf{X} = \mathbf{X}_{(L)}$  to simplify the notations. By calling  $\mathbf{S} \in \mathbb{R}^{L \times N}$  the matrix of binary indicator variables such that :

$$s_{ln} = \begin{cases} 1 & \text{if } l \in C_n, \\ 0 & \text{else} \end{cases} \quad (30)$$

we can rewrite the cost function, Eq. (7) [27], as :

$$\sum_{n=1}^N \sum_{l \in C_n} \|\mathbf{x}_{l:} - \theta_n\|_F^2 = \sum_{n=1}^N \sum_{l \in C_n} s_{ln} \|\mathbf{x}_{l:} - \theta_n\|_F^2, \quad (31)$$

$$= \|\mathbf{X} - \mathbf{S}\Theta\|_F^2, \quad (32)$$

$$= \operatorname{Tr}(\mathbf{X}\mathbf{X}^t) - 2\operatorname{Tr}(\mathbf{S}\Theta\mathbf{X}^t) + \operatorname{Tr}(\mathbf{S}\Theta\Theta^t\mathbf{S}^t), \quad (33)$$

$$= \operatorname{Tr}(\mathbf{X}\mathbf{X}^t) - \operatorname{Tr}(\mathbf{S}\Theta\mathbf{X}^t), \quad (34)$$

The passage from (33) to (34) is proved in the following equations. When  $\mathbf{S}$  is fixed, the centroids are found by least square regression, using the Moore-Penrose pseudoinverse of  $\mathbf{S}$ , noted  $\mathbf{S}^\dagger \in \mathbb{R}^{N \times L}$  (we recall that  $\mathbf{S}^\dagger \mathbf{S} = \mathbf{I}$ ):

$$\Theta = (\mathbf{S}^t \mathbf{S})^{-1} \mathbf{S}^t \mathbf{X} = \mathbf{S}^\dagger \mathbf{X}, \quad (35)$$

Note this result has a physical meaning; since  $\mathbf{S}$  is column orthogonal, we have

$$\theta_n = \frac{\sum_{l=1}^L s_{ln} \mathbf{x}_{l:}}{\sum_{l=1}^L s_{ln}} = \frac{1}{c_n} \sum_{l \in C_n} \mathbf{x}_{l:}. \quad (36)$$

with  $c_n$  the cardinal of cluster  $C_n$ . Henceworth it corresponds to the average of all FCs belonging to the cluster  $C_n$ . Then the first term of Eq. (33) is:

$$\operatorname{Tr}(\mathbf{X}\mathbf{X}^t) = \|\mathbf{X}\|_F^2 = Cste \quad (37)$$

By noticing that  $\mathbf{S}\mathbf{S}^\dagger$  is a projection matrix (and that  $(\mathbf{S}\mathbf{S}^\dagger)^t \mathbf{S}\mathbf{S}^\dagger = \mathbf{S}\mathbf{S}^\dagger$ ), the third term is:

$$\operatorname{Tr}(\mathbf{S}\Theta\Theta^t\mathbf{S}^t) = \operatorname{Tr}((\mathbf{S}\mathbf{S}^\dagger)^t \mathbf{S}\mathbf{S}^\dagger \mathbf{X}\mathbf{X}^t) \quad (38)$$

$$= \operatorname{Tr}(\mathbf{S}\Theta\mathbf{X}^t), \quad (39)$$

This proves the passage from Eq. (33) to (34). Finally:

$$\operatorname{Tr}(\mathbf{S}\Theta\mathbf{X}^t) = \operatorname{Tr}(\mathbf{S}(\mathbf{S}^t \mathbf{S})^{-1} \mathbf{S}\mathbf{X}\mathbf{X}^t), \quad (40)$$

$$= \operatorname{Tr}((\mathbf{S}^t \mathbf{S})^{-\frac{1}{2}} \mathbf{S}^t \mathbf{X}\mathbf{X}^t \mathbf{S}(\mathbf{S}^t \mathbf{S})^{-\frac{1}{2}}), \quad (41)$$

$$= \|\mathbf{A}^t \mathbf{X}\|_F^2, \quad (42)$$

where  $\mathbf{A} \in \mathbb{R}^{L \times N}$  is to the normalised indication matrix,  $\mathbf{A} = \mathbf{S}(\mathbf{S}^t \mathbf{S})^{-\frac{1}{2}}$ , such that:

$$a_{ln} = \begin{cases} \frac{1}{\sqrt{c_n}} & \text{if } l \in C_n, \\ 0 & \text{else} \end{cases}. \quad (43)$$

As  $\|\mathbf{X}\|_F^2$  is fixed, we can see that minimizing Eq. (7) is equivalent to maximizing the second element of Eq. (34). By noticing that columns of  $\mathbf{A}$  form a standard basis of a subspace of dimension  $K$  of  $\mathbb{R}^{L \times L}$ , the optimization problem (7) becomes equivalent to [59]:

$$\underset{\mathbf{A}}{\operatorname{argmax}} \quad \|\mathbf{A}^t \mathbf{X}\|_F^2 \quad (44)$$

$$\text{s.t.} \quad \mathbf{a}_{:,i} \text{ standard basis vector} \quad (45)$$

## APPENDIX B

### MODIFICATION IN THE LOSS FUNCTIONS

#### A. Demonstration of (17) and (16)

We first propose a variant of the cost function (13), using the 3-modes matricization of the tensor  $\mathcal{X}$  (both loss function are equivalent):

$$f(\mathbf{U}, \mathbf{V}, \mathbf{w}) = \|\mathbf{X}_{(S)} - \mathbf{w} \mathbf{G}_{(S)}(\mathbf{V} \otimes \mathbf{U})^t\|_F^2, \quad (46)$$

by reminding  $\mathbf{G}_{(S)} = \mathbf{w}^t \mathbf{X}_{(S)}(\mathbf{V} \otimes \mathbf{U})$  we have:

$$f(\mathbf{U}, \mathbf{V}, \mathbf{w}) = -\operatorname{Tr}(\mathbf{w}^t \mathbf{X}_{(S)}(\mathbf{V} \otimes \mathbf{U})(\mathbf{V} \otimes \mathbf{U})^t \mathbf{X}_{(S)}^t \mathbf{w}) + Ct \quad (47)$$

with  $Ct$  a constant ( here,  $Ct = \operatorname{Tr}(\mathbf{X}_{(S)} \mathbf{X}_{(S)}^t)$ ). By using the variant of the loss function proposed in (47) we can write:

$$f(\mathbf{U}, \mathbf{V}, \mathbf{w}) = -\|\mathbf{w}^t \mathbf{X}_{(S)}(\mathbf{V} \otimes \mathbf{U})\|_F^2 + Ct, \quad (48)$$

Then minimizing  $f$  over the variable  $\mathbf{w}$  with the constraint  $\mathbf{w}^t \mathbf{w} = 1$  corresponds to maximizing the criteria (17). We can also consider that:

$$\operatorname{vec}(\mathbf{X}_L(\mathbf{w} \otimes \mathbf{I}_T))^t = \mathbf{w}^t \mathbf{X}_{(S)}, \quad (49)$$

And the compatibility formula between the vectorization and the Kronecker products [11], which asserts that for three matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{M}$ , we have:

$$\operatorname{vec}(\mathbf{A} \mathbf{M} \mathbf{B}^t)^t = \operatorname{vec}(\mathbf{M})^t (\mathbf{A} \otimes \mathbf{B})^t \quad (50)$$

We can deduce that:

$$f(\mathbf{U}, \mathbf{V}, \mathbf{w}) = -\|\operatorname{vec}(\mathbf{U}^t \mathbf{X}_L(\mathbf{w} \otimes \mathbf{I}_T)^t \mathbf{V})\|_F^2 + Ct, \quad (51)$$

Since we use the Frobenius norm, the  $\operatorname{vec}(\bullet)$  operator is optional. Then minimizing  $f$  over the variable  $\mathbf{U}$  and  $\mathbf{V}$  under semi-orthogonal constraints corresponds to maximizing the criteria (16).

#### B. Demonstration of (20)

We propose the variant of the cost function  $g$  of Eq. (15), using the 3-mode matricization of the tensor  $\mathcal{X}$  (by noticing  $\mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t = \mathbf{F} \mathbf{I}_K(\mathbf{w} \otimes \mathbf{V})^t$ , with  $\mathbf{I}_K \in \mathbb{R}^{K \times K}$  an identity matrix):

$$g(\mathbf{F}, \mathbf{V}, \mathbf{w}) = \|\mathbf{X}_{(S)} - \mathbf{w} \operatorname{vec}(\mathbf{I}_K)^t (\mathbf{V} \otimes \mathbf{F})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \quad (52)$$

Then, if we set  $\mathbf{w} = \mathbf{w}_i$  with  $\mathbf{w}_i^t \mathbf{w}_i = 1$ , we can write:

$$g(\mathbf{F}, \mathbf{V}, \mathbf{w}_i) = \|\mathbf{w}_i^t \mathbf{X}_{(S)} - \operatorname{vec}(\mathbf{I}_K)^t (\mathbf{V} \otimes \mathbf{F})^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \quad (53)$$

Finally, using the compatibility formula between the vectorization and the Kronecker products [11], Eq. (50), and considering Eq. (49), we can write:

$$g(\mathbf{F}, \mathbf{V}, \mathbf{w}_i) = \|\operatorname{vec}(\mathbf{X}_{(S)}(\mathbf{w}_i \otimes \mathbf{I}_T) - \mathbf{F} \mathbf{V}^t)^t\|_F^2 + \lambda_1 \|\mathbf{F}\|_1, \quad (54)$$

Since we use the Frobenius norm, the  $\operatorname{vec}(\bullet)$  operator is optional, then minimizing  $f$  over the variable  $\mathbf{U}$  and  $\mathbf{V}$  corresponds to maximizing the criteria (20).

#### C. Demonstration of (21)

Consider the loss function  $g$  written using 3-mode matricizations, given in Eq. (52). If we set  $\mathbf{F} = \mathbf{F}_{i+1}$  and  $\mathbf{V} = \mathbf{V}_{i+1}$  then we have:

$$g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}) = \|\mathbf{X}_{(S)} - \mathbf{w} \operatorname{vec}(\mathbf{I}_K)^t (\mathbf{V}_{i+1} \otimes \mathbf{F}_{i+1})^t\|_F^2 + Ct, \quad (55)$$

Moreover, considering that  $\mathbf{I}_K$  is a diagonal matrix, we can simplify the computation using the Khatri-Rao product:

$$\operatorname{vec}(\mathbf{I}_K)^t (\mathbf{V} \otimes \mathbf{F})^t = \operatorname{diag}(\mathbf{I}_K)^t (\mathbf{V} \odot \mathbf{F})^t \quad (56)$$

where  $\operatorname{diag}(\mathbf{I}_K) \in \mathbb{R}^K$  the vector representing the main diagonal of  $\mathbf{I}_K$ . Thus we obtain as a loww function:

$$g(\mathbf{F}_{i+1}, \mathbf{V}_{i+1}, \mathbf{w}) = \|\mathbf{X}_{(S)} - \mathbf{w} \operatorname{diag}(\mathbf{I}_K)^t (\mathbf{V} \odot \mathbf{F})^t\|_F^2 + Ct, \quad (57)$$

which, under the constraint  $\mathbf{w}^t \mathbf{w} = 1$ , gets the same minimum as Eq. (21).

#### D. Note on empirical convergence of mHOOI

To empirically observe the convergence towards the optimal solution, with faster performance than HOOI algorithm we propose the following experiment. We consider a tensor  $\mathcal{X}_{model} \in \mathbb{R}^{66 \times 1000 \times 4}$  which corresponds to the model of Section V-A, with a noise vector of  $b = [0.2, 1, 1, 0.3]$ . We fix  $K_L = K_T = 4$  and  $K_S = 1$ . We compare the performance of both HOOI and mHOOI algorithm by computing the total variance at each iteration  $tvar(t) = \|\mathbf{G}_i\|_F^2$  for HOOI,  $tvar(t) = \|\mathbf{F}_i\|_F^2$  (since they are matrices containing all the variance) and the log differential of the total variance between two consecutive iterates  $ldvar(t) = \log(tvar(t) - tvar(t-1))$ . Fig. 7 shows the mean of  $tvar(t)$  for the first 100 iterations using 100 realisations of the tensor  $\mathcal{X}_{model}$  as input of the algorithm.



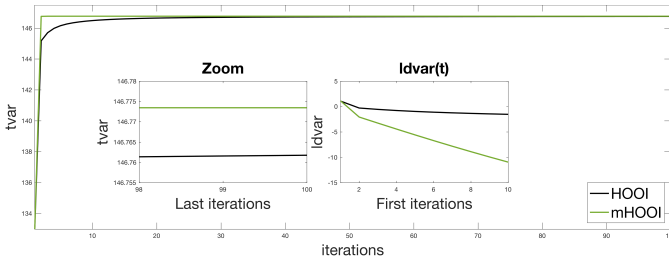


Fig. 7:  $tvar(t)$  for the first 100 iterations using the tensor  $\mathcal{X}_{model}$  as input of the algorithm with a zoom of the last iteration and the first value of  $ldvar(t)$

A zoom of the last iteration is provided as well as the first value of  $ldvar(t)$ . The new proposed algorithm has always better performances than HOOI, with exponential convergence, on this scenario. This experiment, tested for other configurations, gave similar results.

## APPENDIX C

### DIMENSIONALITY REDUCTION WITH NMF AND NNCP

Non-negativity constraints for non-negative data was shown to be useful in a lot of applications. Performing the decomposition  $\mathbf{X}_{(L)} \approx \mathbf{F}\mathbf{Z}^t$ , with  $\mathbf{Z} \in \mathbb{R}^{T \times S \times K}$  and imposing  $\mathbf{f}_{lk} \geq 0$  and  $\mathbf{z}_{tk} \geq 0$ , leading thus to the so-called non-negative matrix factorisation (NMF) [60]. NMF is also viewed as a soft clustering method [61], but its use for dimensionality reduction prior to  $k$ -means clustering is less common. Still, we will compare the use of the corresponding low dimension factor matrix  $\mathbf{F}_{NMF} = \mathbf{F}$  in Eq. (9) and check its performance against other alternatives. We normalise the columns of  $\mathbf{F}$  and of  $\mathbf{Z}$  such that  $\|\mathbf{z}_{:k}\|_F^2 = 1 \forall k \in 1, \dots, K$ , in order to remove the scaling indeterminacies between both matrices.

The most popular tensor decomposition is certainly the canonical polyadic decomposition [62], which extends the SVD in the sense that it decomposes the data into a sum of rank one tensors, like SVD decomposes the data in a sum of rank one matrices. The decomposition can be noted  $\mathbf{X}_{(L)} \approx \mathbf{F}(\mathbf{W} \odot \mathbf{V})^t$ . To warrant the existence of a global solution for this decomposition, it is recommended to add non-negativity constraints to all factors [63], leading to the non-negative canonical polyadic decomposition [64] (nnCP). After rescaling  $\mathbf{U}$  such that  $\|\mathbf{v}_{:k}\|_F^2 = \|\mathbf{w}_{:k}\|_F^2 = 1 \forall k \in 1, \dots, K$  in order to avoid scaling indeterminacies, we get  $\mathbf{F}_{nnCP} = \mathbf{U}$  to be substituted in Eq. (9).

## REFERENCES

- [1] M. Guenot, J. Isnard, P. Ryvlin, C. Fischer, K. Ostrowsky, F. Mauguire, and M. Sindou, "Neurophysiological monitoring for epilepsy surgery: the talairach seeg method," *Stereotactic and functional neurosurgery*, vol. 77, no. 1-4, pp. 29–32, 2001.
- [2] P. Chauvel, S. Rheims, A. McGonigal, and P. Kahane, "French guidelines on stereoelectroencephalography (seeg): Editorial comment," *Neurophysiologie clinique= Clinical neurophysiology*, vol. 48, no. 1, p. 1, 2018.

- [3] P. van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, and D. Marinazzo, "Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization," *Progress in neurobiology*, vol. 121, pp. 19–35, 2014.
- [4] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Phys. Rev. E*, vol. 89, p. 032804, Mar 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevE.89.032804>
- [5] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [6] G. Ortiz-Jiménez, M. Coutino, S. P. Chepuri, and G. Leus, "Sampling and reconstruction of signals on product graphs," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 713–717.
- [7] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3675–3687, July 2017.
- [8] F. Sheikholeslami and G. B. Giannakis, "Overlapping community detection via constrained parafac: A divide and conquer approach," in *2017 IEEE International Conference on Data Mining (ICDM)*, Nov 2017, pp. 127–136.
- [9] P. Comon, "Tensors: a brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, 2014.
- [10] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [11] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [12] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [13] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75 – 174, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0370157309002841>
- [14] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1 – 44, 2016, community detection in networks: A user guide. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0370157316302964>
- [15] E. Abbe, "Community detection and stochastic block models," *Found. Trends Commun. Inf. Theory*, vol. 14, no. 1-2, pp. 1–162, Jun. 2018. [Online]. Available: <https://doi.org/10.1561/01000000067>
- [16] P. J. Bickel and A. Chen, "A nonparametric view of network models and newman–girvan and other modularities," *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, pp. 21 068–21 073, 2009. [Online]. Available: <https://www.pnas.org/content/106/50/21068>
- [17] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks," *Phys. Rev. E*, vol. 83, p. 016107, Jan 2011. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevE.83.016107>
- [18] A. N. Khambhati, K. A. Davis, B. S. Oommen, S. H. Chen, T. H. Lucas, B. Litt, and D. S. Bassett, "Dynamic network drivers of seizure generation, propagation and termination in human neocortical epilepsy," *PLoS computational biology*, vol. 11, no. 12, p. e1004608, 2015.
- [19] A. Hegde, D. Erdogmus, and J. C. Principe, "Spatio-temporal clustering of epileptic ecog," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006, pp. 4199–4202.
- [20] J. Gonzalez-Castillo, C. W. Hoy, D. A. Handwerker, M. E. Robinson, L. C. Buchanan, Z. S. Saad, and P. A. Bandettini, "Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns," *Proceedings of the National Academy of Sciences*, vol. 112, no. 28, pp. 8762–8767, 2015.
- [21] F. Liu, Y. Wang, M. Li, W. Wang, R. Li, Z. Zhang, G. Lu, and H. Chen, "Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic-clonic seizure," *Human brain mapping*, vol. 38, no. 2, pp. 957–973, 2017.
- [22] L. Geerligs, N. M. Maurits, R. J. Renken, and M. M. Lorist, "Reduced specificity of functional connectivity in the aging brain

- during task performance,” *Human brain mapping*, vol. 35, no. 1, pp. 319–330, 2014.
- [23] X. Shi, W. Wang, and C. Zhang, “An Empirical Comparison of Latest Data Clustering Algorithms with State-of-the-Art,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 5, no. 2, pp. 410–415, 2017.
  - [24] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Kdd*, vol. 96, 1996, pp. 226–231.
  - [25] B. J. Frey and D. Dueck, “Clustering by passing messages between data points,” *science*, vol. 315, no. 5814, pp. 972–976, 2007.
  - [26] L. R. Chai, A. N. Khambhati, R. Ciric, T. M. Moore, R. C. Gur, R. E. Gur, T. D. Satterthwaite, and D. S. Bassett, “Evolution of brain network dynamics in neurodevelopment,” *Network Neuroscience*, vol. 1, no. 1, pp. 14–30, 2017.
  - [27] C. Bauckhage, “K-means clustering is matrix factorization,” *arXiv preprint arXiv:1512.07548*, 2015.
  - [28] C. Ding and X. He, “K-means clustering via principal component analysis,” in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 29.
  - [29] K. Nur’aini, I. Najahaty, L. Hidayati, H. Murfi, and S. Nurrohman, “Combination of singular value decomposition and K-means clustering methods for topic detection on Twitter,” in *Advanced Computer Science and Information Systems (ICACSIS), 2015 International Conference on*. IEEE, 2015, pp. 123–128.
  - [30] A. Ben-Hur and I. Guyon, “Detecting stable clusters using principal component analysis,” in *Functional genomics*. Springer, 2003, pp. 159–182.
  - [31] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, “Singular value decomposition and principal component analysis,” in *A practical approach to microarray data analysis*. Springer, 2003, pp. 91–109.
  - [32] A. H. Hossny, T. Moschuo, G. Osborne, L. Mitchell, and N. Lothian, “Enhancing keyword correlation for event detection in social networks using SVD and k-means: Twitter case study,” *Social Network Analysis and Mining*, vol. 8, no. 1, p. 49, 2018.
  - [33] F. Pisana, T. Henzler, S. Schönborg, E. Klotz, B. Schmidt, and M. Kachelrieß, “High quality high spatial resolution functional classification in low dose dynamic CT perfusion using singular value decomposition (SVD) and k-means clustering,” in *Medical Imaging 2017: Physics of Medical Imaging*, vol. 10132. International Society for Optics and Photonics, 2017, p. 101320M.
  - [34] A. A. Shabalin and A. B. Nobel, “Reconstruction of a low-rank matrix in the presence of Gaussian noise,” *Journal of Multivariate Analysis*, vol. 118, pp. 67–76, 2013.
  - [35] D. Feldman, M. Schmidt, and C. Sohler, “Turning big data into tiny data: Constant-size coresets for k-means, pca and projective clustering,” in *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2013, pp. 1434–1453.
  - [36] A. Cichocki, “Tensor decompositions: a new concept in brain data analysis?” *arXiv preprint arXiv:1305.0395*, 2013.
  - [37] A. Zhang and D. Xia, “Tensor SVD: Statistical and Computational Limits,” *IEEE Transactions on Information Theory*, 2018.
  - [38] L. De Lathauwer, B. De Moor, and J. Vandewalle, “On the best rank-1 and rank-( $r_1, r_2, \dots, r_n$ ) approximation of higher-order tensors,” *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.
  - [39] Y. Xu, “On the convergence of higher-order orthogonal iteration,” *Linear and Multilinear Algebra*, vol. 66, no. 11, pp. 2247–2265, 2018.
  - [40] A. G. Mahyari, D. M. Zoltowski, E. M. Bernat, and S. Aviyente, “A tensor decomposition-based approach for detecting dynamic network states from eeg,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 225–237, 2017.
  - [41] A. Ozdemir, E. M. Bernat, and S. Aviyente, “Recursive tensor subspace tracking for dynamic brain network analysis,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 4, pp. 669–682, 2017.
  - [42] M. J. Tobia, K. Hayashi, G. Ballard, I. H. Gotlib, and C. E. Waugh, “Dynamic functional connectivity and individual differences in emotions during social stress,” *Human brain mapping*, vol. 38, no. 12, pp. 6185–6205, 2017.
  - [43] F. Bach, J. Mairal, and J. Ponce, “Convex sparse matrix factorizations,” *arXiv preprint arXiv:0812.1869*, 2008.
  - [44] G. Frusque, J. Jung, P. Borgnat, and P. Gonçalves, “Sparse tensor dimensionality reduction with application to clustering of functional connectivity,” 2019.
  - [45] L. De Lathauwer, “Decompositions of a higher-order tensor in block terms—part ii: Definitions and uniqueness,” *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, 2008.
  - [46] M. Mørup, L. K. Hansen, and S. M. Arnfred, “Algorithms for sparse nonnegative tucker decompositions,” *Neural computation*, vol. 20, no. 8, pp. 2112–2131, 2008.
  - [47] Z. Lai, Y. Xu, Q. Chen, J. Yang, and D. Zhang, “Multilinear sparse principal component analysis,” *IEEE transactions on neural networks and learning systems*, vol. 25, no. 10, pp. 1942–1950, 2014.
  - [48] G. Allen, “Sparse higher-order principal components analysis,” in *Artificial Intelligence and Statistics*, 2012, pp. 27–36.
  - [49] R. Badeau and R. Boyer, “Fast multilinear singular value decomposition for structured tensors,” *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1008–1021, 2008.
  - [50] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online learning for matrix factorization and sparse coding,” *Journal of Machine Learning Research*, vol. 11, no. Jan, pp. 19–60, 2010.
  - [51] D. Yang, Z. Ma, and A. Buja, “A sparse singular value decomposition method for high-dimensional data,” *Journal of Computational and Graphical Statistics*, vol. 23, no. 4, pp. 923–942, 2014.
  - [52] P. Stoica and G. Ganesan, “Linear regression constrained to a ball,” *Digital Signal Processing*, vol. 11, no. 1, pp. 80–90, 2001.
  - [53] C. A. Andersson and R. Bro, “The N-way toolbox for MATLAB,” *Chemometrics and intelligent laboratory systems*, vol. 52, no. 1, pp. 1–4, 2000.
  - [54] B. W. Bader and T. G. Kolda, “Matlab tensor toolbox version 2.5,” *Available online, January*, vol. 7, 2012.
  - [55] J. Mairal, F. Bach, J. Ponce, G. Sapiro, R. Jenatton, and G. Obozinski, “SPAMS: A SParse Modeling Software, v2. 3,” *URL <http://spams-devel.gforge.inria.fr/downloads.html>*, 2014.
  - [56] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.
  - [57] L. Hubert and P. Arabie, “Comparing partitions,” *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.
  - [58] F. Wendling, F. Bartolomei, J.-J. Bellanger, J. Bourien, and P. Chauvel, “Epileptic fast intracerebral eeg activity: evidence for spatial decorrelation at seizure onset,” *Brain*, vol. 126, no. 6, pp. 1449–1459, 2003.
  - [59] J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine learning refined: foundations, algorithms, and applications*. Cambridge University Press, 2016.
  - [60] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Advances in neural information processing systems*, 2001, pp. 556–562.
  - [61] C. C. Aggarwal and C. K. Reddy, *Data clustering: algorithms and applications*. CRC press, 2013.
  - [62] R. Bro, “PARAFAC. Tutorial and applications,” *Chemometrics and intelligent laboratory systems*, vol. 38, no. 2, pp. 149–171, 1997.
  - [63] L.-H. Lim and P. Comon, “Nonnegative approximations of nonnegative tensors,” *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 23, no. 7-8, pp. 432–441, 2009.
  - [64] Y.-D. Kim and S. Choi, “Nonnegative tucker decomposition,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.